2016

Ribozyme/Duplex Binding Interactions as a Thermodynamic Basis for Chemical Game Theory

Elizabeth Satterwhite
Portland State University

Ribozyme/Duplex Binding Interactions as a Thermodynamic Basis for Chemical Game Theory

Elizabeth Satterwhite
Dr. Niles Lehman, Research Advisor
Portland State University
Department of Chemistry
June 2016

Thesis submitted in partial satisfaction of the requirements for the completion of a Bachelor's of Science Honors degree in Biochemistry

Table of Contents

Abstract

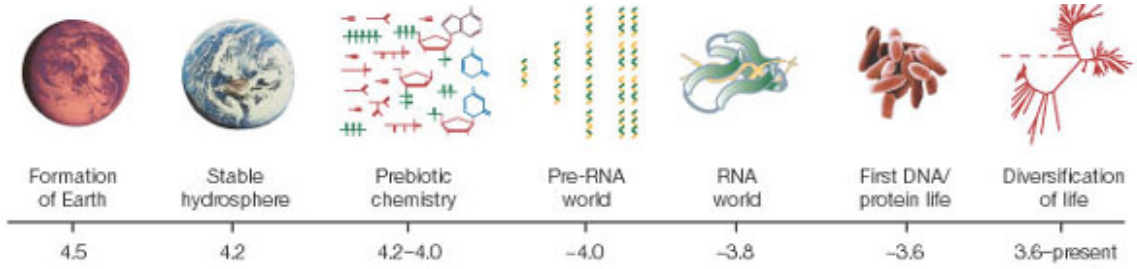
Origins of life research requires searching for a plausible transition from simple chemicals to larger macromolecules that can both hold information and catalyze their own production. It has been shown that some classes of RNA molecules possess the ability to help other RNA molecules form by recombination reactions in an auto-catalytic fashion.[1] By simplifying these recombination reactions, the thermodynamic binding strength between two nucleotides of two molecules can be quantified for all 16 possible genotype combinations. The purpose of this research is to provide evidence that the thermodynamic binding strength between the tag and internal guide sequence (IGS) is correlated to the catalytic ability of the ribozyme, or the specific relationship between various tag and IGS base pair combinations.

Introduction

Interest in the origins of life on Earth has existed since humans were first able to ponder such a daunting question. Revelations within the past century on the structure of DNA (deoxyribonucleic acid), the central dogma of molecular biology and the catalytic ability of RNA (ribonucleic acid) has lead scientists to new insights as to what the first life on primitive Earth may have looked like. Hypothesizing what the first life on Earth looked like cannot occur without first setting standards for what constitutes 'life'. NASA defines life rather broadly as: 'A self-sustaining chemical system capable of…evolution.'[2] Evolution can generally be regarded as gradual change over time, usually to a more complex system from a simpler one, in which the gradual change is fueled by variety in the individuals subject to evolution. Thus, life can only be defined as such if the chemicals can sustain their own reproduction, and the information they hold can evolve over time.

The central dogma of molecular biology states that the genetic information contained within DNA is phenotypically expressed by first transcription of a particular section into messenger RNA, followed by translation into a protein. This transfer of information from DNA to RNA to proteins is how most contemporary organisms express their genes. In order to study origins of life hypotheses, scientists much consider a time before the central dogma of molecular biology applied, or a time in which DNA, RNA, or proteins did not yet exist. Figure 1 depicts a hypothesized timeline of the genesis of life on Earth. This figure suggests that prior to the emergence of life, prebiotic chemicals existed on Earth, which then evolved into primitive molecules, giving rise to a period of time known as the 'RNA World'.

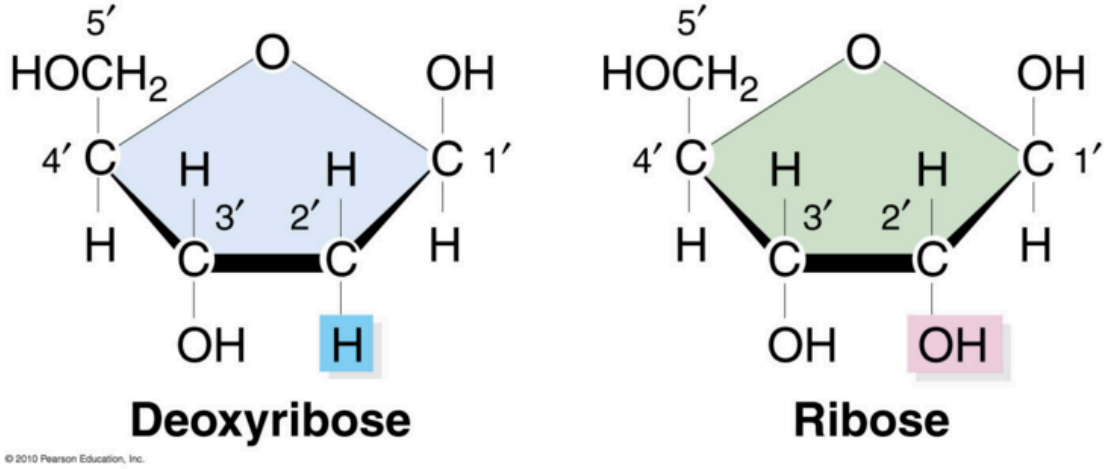
Figure 1: Timeline of the emergence of life on primitive Earth.[3]



The RNA World Hypothesis suggests that Earth's history includes a period of time in which RNA or RNA-like molecules existed before DNA and protein molecules.

This method of thought solves a 'chicken or the egg' dilemma when attempting to determine which of the three information-containing molecules arose first. Between the options of DNA, RNA, or proteins, RNA is the only one able to both store genetic information and catalyze its own reproduction. DNA can store information but cannot catalyze reactions. Proteins can potentially catalyze reactions but cannot store the genetic information that dictates their structure. The catalytic ability of RNA lies within a single oxygen atom located at the second carbon (C2') of the ribose sugar, which partially constitutes both DNA and RNA monomers. Figure 2 displays the subtle but imperative difference between individual units of DNA and RNA.

Figure 2: ribose sugar constituents of DNA and RNA monomers.[4]



In order for the RNA World Hypothesis to be a plausible scenario as to how life began on Earth, scientists must be able to show that RNA molecules are living, meaning they must be able to form a self-sustaining chemical system capable of evolution. Thus it must be shown that catalytic RNA molecules (ribozymes) can independently reproduce, while adapting to changes in the environment and changes in genome arising from error ultimately driving evolution towards greater complexity. Previous experiments have demonstrated how RNA molecules can interact in a primitive reproduction-like manner.[1] In these experiments, a group I intron (class of ribozyme) from the *Azoarcus* bacterium was split into four pieces denoted 'W', 'X', 'Y', and 'Z'. These fragments were then observed to bind together by recombination reactions to form the full-length molecule. This full-length molecule then assists in the catalysis of other recombination reactions,

exponentially increasing the rate of full molecule formation in an autocatalytic fashion (Figure 3).

Figure 3: Recombination of fragments to form full ribozyme.

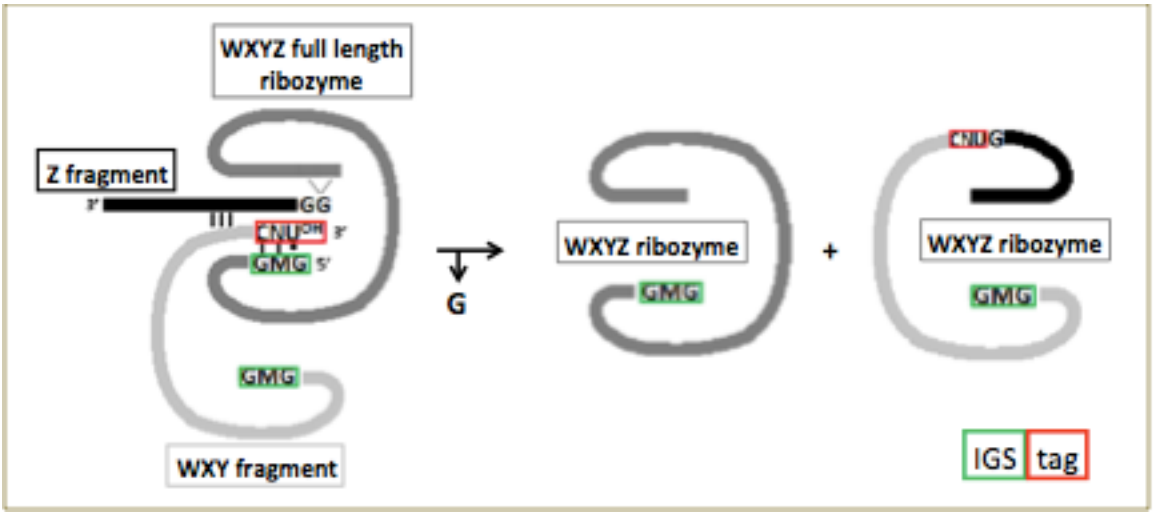


Figure 3 shows the recombination of WXY (previously recombined from W, X, and Y) and Z. In the reaction, the recombination of two fragments of the ribozyme requires recognition of a 5' IGS (located on the full ribozyme directing the recombination) to a 3' tag sequence (located on one of the fragments being combined). Both the IGS and tag are comprised of three nucleotides; the 'WXY' fragment contains a tag of 'C**N**U' that binds to the fully formed ribozyme's IGS of 'G**M**G'. The identity of the 'M' and 'N' can be any of the four canonical RNA nucleotides: adenine (A), cytosine (C), guanine (G), or uracil (U). The identification of 'M' and 'N' in part determines how well the recombination reaction occurs.

The experiments described above provide evidence that the RNA World Hypothesis can meet the criteria of a self-sustaining chemical system, but the molecules must also be capable of evolution. When considering the RNA World, the notion of evolution implies that there must be variation in genotypes of the molecules present to make survival and evolution to the diversity of life seen today more likely. Scientists must also be able to show that the same RNA molecules varying only in genotype (the three nucleotide IGS) can interact and assist one another in replication. Previous experiments have shown that cooperative networks of the same *Azoarcus* intron in which one genotype assembles another in a cyclic fashion causes RNA populations to evolve

greater complexity as compared to the previously described 'selfish' autocatalytic replication.[5]

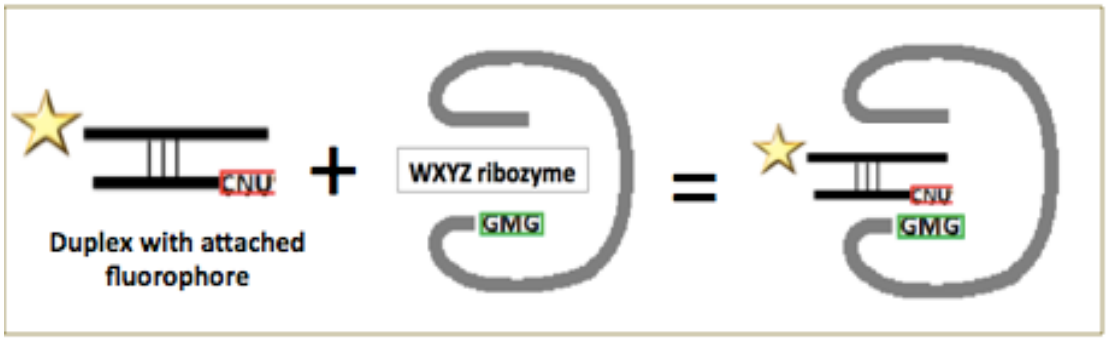
Having shown that a class of catalytically active RNA molecules meet the broad definition of life, questions still remain as to what a population of ribozymes in the RNA World might have looked like. In a system with individuals of different genotypes, each genotype will compete to pass its information on to the next generation as is seen in contemporary organisms today. The study of population frequencies as a result of the presence of other individuals brings about the notion of game theory. Game theory is generally defined as "the study of mathematical models of conflict and cooperation between intelligent rational decision-makers."[6] Game theory is a useful model in many disciplines, including biology to analyze population outcomes based on competition for resources, or a 'game' between different organisms or 'players'. On a molecular and chemical level, 'players' can refer to the ribozymes and their catalytic ability, with competition for survival of a genotype by assembly of one's genotype becoming the 'game'. When applying game theory concepts to organic molecules with catalytic activity such as these, it can be shown that 'intelligent rational decision-makers' are not necessarily required to comprise a game theoretic system.

A game theoretic analysis of the autocatalytic recombination system with varying genotypes within the IGS and tag demonstrates different catalytic abilities of the ribozymes dependent upon the presence and frequency of other genotypes in the system. While this catalytic activity can be measured, it is hypothesized that the underlying thermodynamic basis for the success of one genotype over another, or 'fitness' resides in the internal guide sequence/tag binding interaction that lines up the catalytic site. In order to address this possibility, we have tested the binding ability of each possible IGS/tag combination. It is hypothesized that there is a correlation between the genotypes and phenotypes of the ribozymes (i.e. the 'MN' combination and catalytic activity) and the thermodynamic strength of the bond. Finding evidence for this correlation will help to define the contribution of the individual nucleotides, or primary tag/IGS combination to the overall tertiary structure interactions and catalytic abilities of the *Azoarcus* group I intron.

Methods

In order to be able to quantitatively determine the binding strength of each nucleotide pair combination, the recombination reaction cannot be allowed to take place. The equilibrium constants of an autocatalytic recombination reaction cannot be determined by a classic Michaelis-Menten analysis, since the fragments that are recombined by a ribozyme becomes another ribozyme for other recombination reactions. This autocatalytic process introduces more complexity into the concentration of ribozyme at any one point, making determining the strength of the 'MN' bond using the *in vivo* reaction quite difficult.[7] Hence, a modified but similar reaction was used to determine the binding strengths in terms of the dissociation constant (K_d), and ultimately the Gibbs free energy. A duplex was designed to mimic the 'WXY' and 'Z' fragments that are normally recombined by the fully formed ribozyme (Figure 4).

Figure 4: Modified 'recombination' reaction scheme. Note the similarity of this scheme to that presented in Figure 3.



This duplex was carefully designed so as to affect the reaction dynamics as little as possible. Therefore, the sequence of nucleotides in the duplex binding region are the same as in the 'WXY' and 'Z' fragments, except for a few replacements of nucleotides with C and G. C and G were selectively added, as a higher frequency of these nucleotides increases the stability of a molecule at higher temperatures, such as the reaction conditions for these experiments. The duplex was also specifically designed to be much smaller than the ribozyme, so that a noticeable change in polarization could be observed upon binding. Also in the modified reaction, when the duplex's tag binds to the ribozyme's IGS, recombination does not occur because the 3' end of the tag sequence (U of CNU) is intentionally dehydroxylated. In reference to Figure 2, removal of the oxygen from the ribose portion of the U nucleotide ensures that there is no catalytic activity, and

therefore recombination is not possible. With this modified scheme, the binding strength between the middle nucleotides of the internal guide sequence and tag can be determined for all 16 possible combinations. 6-FAM (a fluorophore) was added to the 5' end of the 'Z' fragment of the duplex in order to quantify the ratio of bound to free ribozyme when using fluorescence anisotropy.

Fluorescence anisotropy is a well-suited tool for studying molecular binding interactions. By observing changes in the size of a fluorescently labeled molecule (the duplex binding to the ribozyme) the strength of the bond and Gibbs free energy can ultimately be determined. Polarized light is used to determine when the duplex is binding to the ribozyme. For example, if the duplex (24 nt) is not bound to the ribozyme (200 nt) in solution, it will move around very quickly as a smaller molecule; as a result, light traveling through the cuvette will be depolarized (scattered in various directions) by the small, rapidly moving molecule. However, if the duplex is bound to the ribozyme, the bound molecule being larger tumbles much more slowly in solution, allowing polarized light to pass through the cuvette, thus giving a higher polarization value. Because of the methodology by which FA works, it was notably important for the duplex to be significantly smaller than the ribozyme, so that a noticeable change in polarization will occur upon binding (this is not the case in the standard reaction, where sum of the fragments are the same length as the full ribozyme).

Experimental Procedures

The 'Y' and 'Z' fragments of the duplex ('Y' in the modified reaction is synonymous to 'WXY' in the standard reaction, but much shorter) were ordered separately from Integrated DNA Technologies (IDT). 'Y' and 'Z' were annealed by heating the fragments in solution together to 80 °C, and then allowing them to cool to 20 °C over the course of several hours. All FA experiments were performed at 48 °C. The reaction buffer used in these FA experiments was made from a 5x reaction buffer stock made of 500mM Mg^{2+} and 150mM EPPS at pH 7.5, and was diluted to 1x as needed for experiments. These reaction conditions were chosen to match the conditions under which the *in vivo* self-assembly recombination reactions were observed. Fluorescence anisotropy experiments were performed on a PerkinElmer L5 SS Fluorescence

Spectrometer using FL Win Lab, with the temperature maintained by a PE Temperature Programmer. 150 μL 1x reaction buffer is dispensed into cuvette, allowed to heat, and 1 μL duplex is added. Ribozyme is then added in 1-2 μL increments, with at least four polarization values recorded for each addition, the average of which is used for graphing results. Binding plots were made with Kaleidagraph software.

All fully formed ribozymes used in these experiments were synthesized in a molecular evolution laboratory, with most prepared by *in vitro* transcription from DNA plasmid templates. Some pre-transcription DNA was synthesized using VATR (vent-assisted template reconstruction). All RNAs were gel purified prior to FA experiments. To provide a control, a ribozyme was also produced that contains no IGS sequence to observe in solution with a duplex using FA.

Three native gels were also run to explore the possibility of ribozyme aggregation at high concentrations of ribozyme in solution (> 10 μM). These gels were ran at 42 °C and 40 Watts in native 1x TBE buffer. Some native gel samples were mixed with a 5x reaction buffer consisting of 500 mM Mg^{2+} and 150 mM EPPS; the same reaction buffer used in FA experiments. Other native gel samples were mixed with a reduced concentration magnesium reaction buffer consisting of 25 mM Mg^{2+} and 150 mM EPPS, to see if a reduction in magnesium would prevent aggregation at higher ribozyme concentrations.

Results & Discussion

Table 5: Experimental rate constants and predicted Gibbs free energy of binding for recombination interactions[8]

Genotype	k_a (min^{-1})	$\Delta G°_{48}$ predicted (kcal/mol)
CG	0.0415	-4.58
AU	0.0319	-3.00
UA	0.0197	-3.13
GC	0.0125	-5.09
GU	0.0091	-2.23
AC	0.0069	5.48
UG	0.0049	0.34
UC	0.0038	5.48
UU	0.0022	5.37
CA	0.0020	4.75
CC	0.0016	5.48
GG	0.0006	2.79
GA	0.0005	5.48
AA	0.0004	4.86
CU	0.0004	5.27
AG	0.0001	4.03
IGS-less		

Table 5 shows the experimentally determined rate constants and change in Gibbs free energy upon binding of the WXY and Z fragments whose 'M' and 'N' are denoted in the left column. It is important to note that a lowercase k_a as seen in Table 5 indicates a rate constant, or the turnover number of a reaction, while an uppercase K_a or K_d indicates an association or dissociation constant respectively, associated with the equilibrium of a reaction. Thus, the extent of the correlation between the 'MN' combination and catalytic activity with the thermodynamic strength of the bond cannot be inferred from comparison between the predicted rate constants in Table 5 and the experimentally determined K_a values in Table 14. However, if K_a follows the same general decreasing pattern with

poorer binding combinations that is seen in k_a values in Table 5, it can be concluded that the rate constant values have a significant impact on the equilibrium constant values K_a and K_d. If experimental K_a values do not follow the general decreasing pattern, this will indicate that other molecular binding interactions aside from the rate constant are significantly affecting the equilibrium value of the reaction.

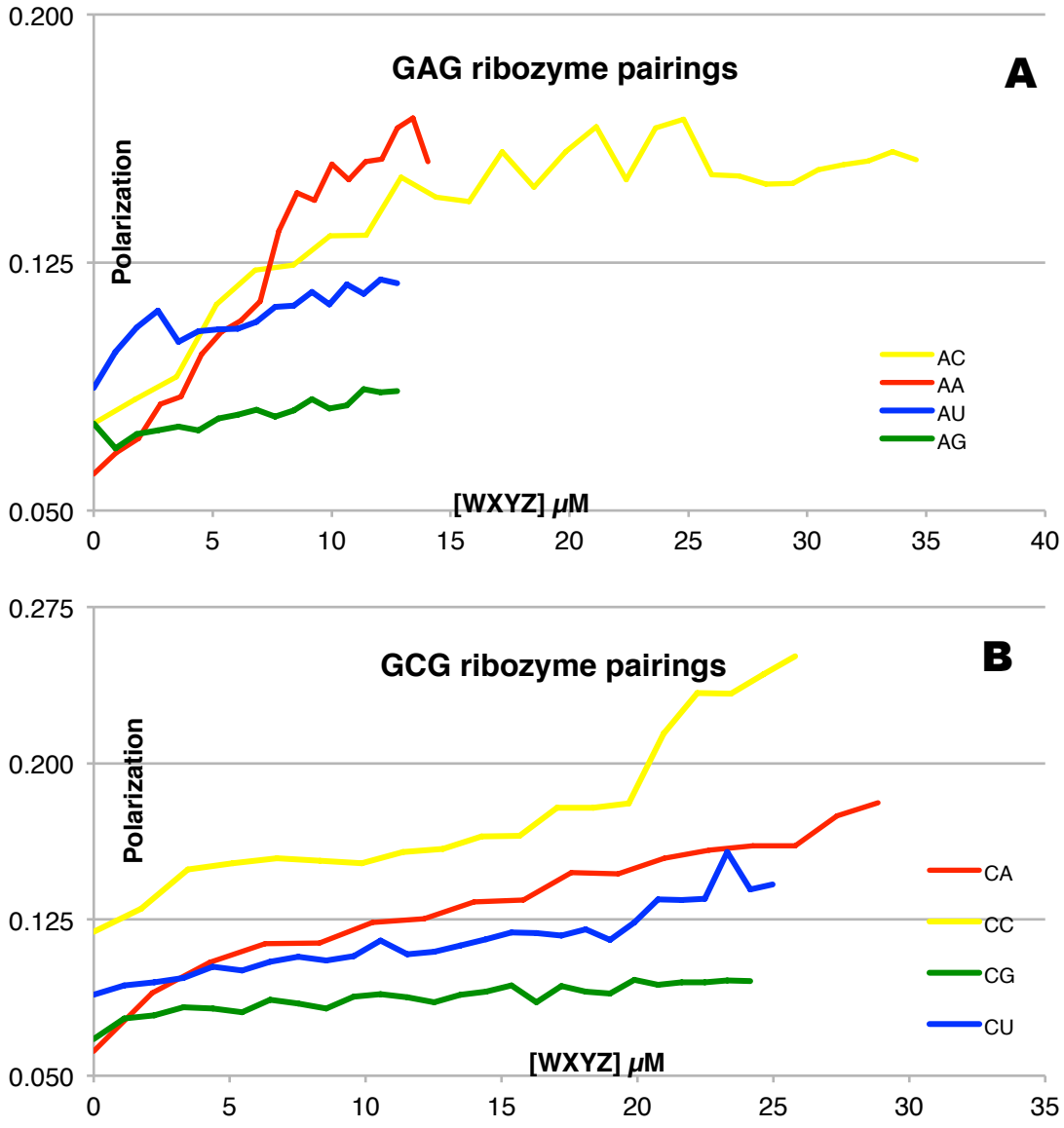
The autocatalytic rate constant prediction values in Table 5 are based on a simplification of the recombination system and thus do not take various binding efficiency details into account.[9, 10] The predicted Gibbs free energy values in Table 5 also represent a gross simplification of the recombination system, as the values only account for the energy of association between the three-nucleotide IGS/tag bond. Since it has been shown that the *Azoarcus* ribozyme has a significantly greater amount of tertiary binding energy than other group I introns of similar size, a Gibbs free energy determined from only the local active site could very well be an inaccurate representation of the overall binding abilities of the ribozyme.[11] Since the experimental data gathered are equilibrium constants, the Gibbs free energy associated with each bond can be directly calculated from K_d by the relation $\Delta G° = -RT\ln K_d$.

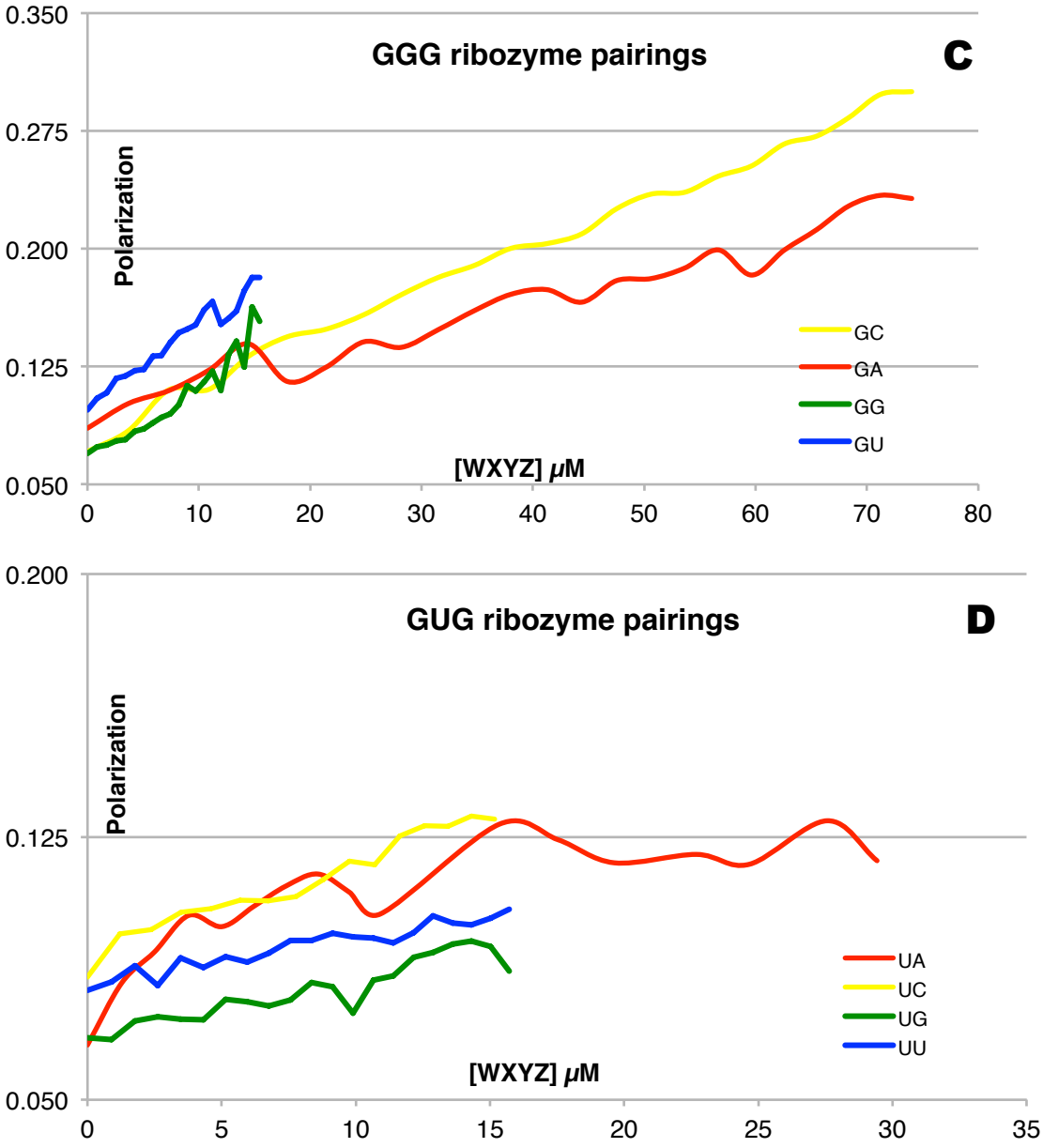
The preliminary data gathered from the fluorescence anisotropy experiments is shown in Figure 6a-d. These plots reflect the increase in polarization as concentration of ribozyme of a particular genotype increases. Ideally, a leveling-off should be seen in which polarization ceases to increase significantly while concentration is still being increased. This leveling-off would indicate a saturation of the duplex, meaning that every duplex molecule is bound to a ribozyme by the IGS/tag binding interaction. In reality, this leveling off is not always seen, particularly with poorer non-canonical/non-wobble nucleotide combinations. The poorer a base pairing is thermodynamically, the higher the required theoretical concentration to bind all duplex molecules.

It is hypothesized that Watson-Crick pairs (CG, GC, AU, UA) would exhibit plots with the most distinct leveling-off in Figure 6a-d. This is because these binding pairs are the most thermodynamically favorable, and thus would most easily achieve a scenario in which all duplex molecules are bound to ribozymes. While this can be seen in both UA, AU and CG, no such behavior is observed for GC. Note that in Figure 6c, the genotypes

'GC' and 'GA' were titrated to at least twice the concentration of other combinations. This extension was done in an effort to determine if a leveling-off could be achieved at a

Figure 6a-d: Ribozyme/duplex pairings





higher concentration. It is hypothesized that at supersaturated concentrations the GGG ribozyme can bind to itself, causing self-aggregation of the ribozyme. This is thought to occur between the IGS and the junctions of another fully formed ribozyme where the fragments comprising the ribozyme were previously recombined. These loop regions include sequences of CCC at the WX junction and CCU at the XY junction, both of which could potentially bind favorably with an IGS of GGG. This non-specific binding would explain the polarization value continuing to increase as ribozyme concentration increases rather than leveling off, as an aggregation of ribozyme molecules would move

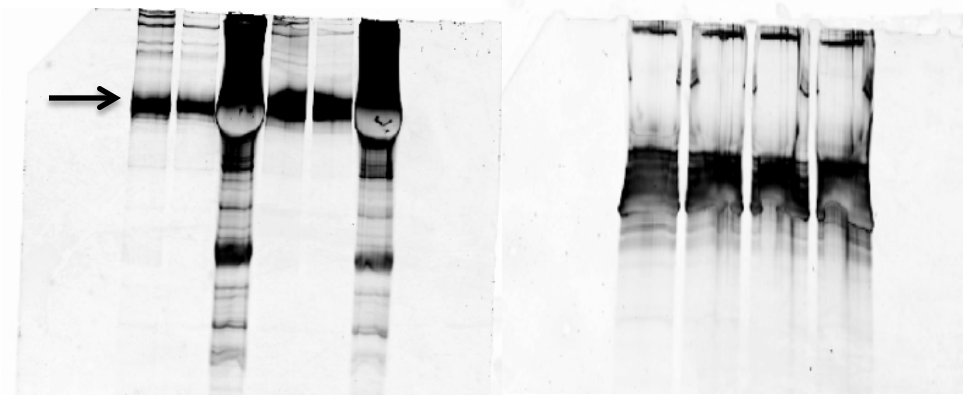
even more slowly in solution than the ribozyme/duplex interaction, allowing more polarized light through the sample.

To test this hypothesis, native gels were run of both the GGG and GUG ribozymes, to test for aggregation at supersaturated concentrations. Note that aggregation will only occur when the concentration of ribozyme is much greater than the concentration of duplex. Since the duplex concentration for all FA experiments is about 0.04 μM, the reaction solution can be thought of as supersaturated once the ribozyme concentration is around 10 μM. Thus it is hypothesized that bands larger than the single ribozyme (200 nt) will be seen on a native gel around ribozyme concentrations of 10 μM or greater. GUG was run as a control, since no non-specific binding should occur for other genotypes besides GGG.

Figures 7a-b show the results of GGG and GUG at 100 mM Mg^{2+} reaction buffer. Figure 7a shows the main 200 nt band (see arrow), as well as bands above these denoting a greater number of nucleotides and thus an aggregation of ribozyme at as little as 6.7 μM. From left to right, the first three lanes in 7a were prepared without a duplex present, while lanes 4-6 were prepared with a duplex. The presence or absence of duplex seems to have little effect on aggregation. Likewise in Figure 7b the left two lanes are samples without duplex, while the right two lanes are with duplex present. Unlike GGG however, the GUG gel does not display the same aggregate banding pattern seen in 7a. The same small dark band seen above each lane in 7b despite the changing concentration is likely a remnant of sample traveling through the gel, which is common at such high magnesium reaction conditions.

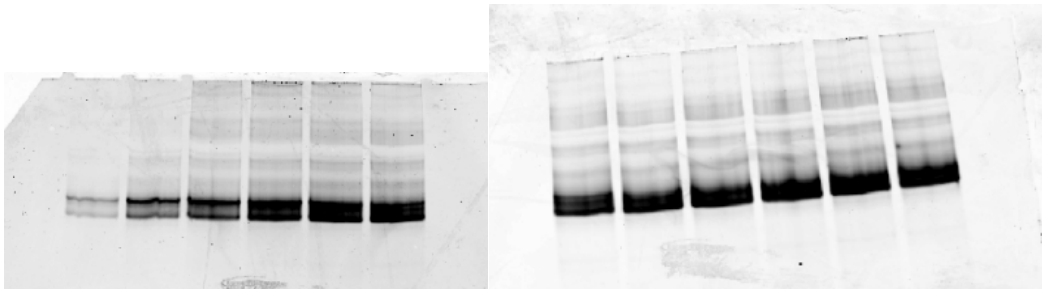
Next, native gels were run with a reaction buffer with a 20-fold reduction in magnesium concentration, to observe whether this reduction would reduce or stop aggregation. Figure 8a-b shows the results of this experiment, where no aggregation bands can be seen even above concentrations of 10 μM. In light of this discovery, one FA experiment with a GGG ribozyme (GC) was re-performed using 5 mM Mg^{2+} in the reaction buffer. These results are shown in Figure 12e. Since aggregation only seems to affect the GGG genotype, all FA experiments were ultimately carried out with 100 mM Mg^{2+} reaction buffer.

Figure 7a-b: Native gels of GGG(a, left) and GUG(b, right) at 100 mM 1x reaction buffer



µM 13.4 6.7 373 12.4 6.2 320 14.4 7.2 6.6 13.2

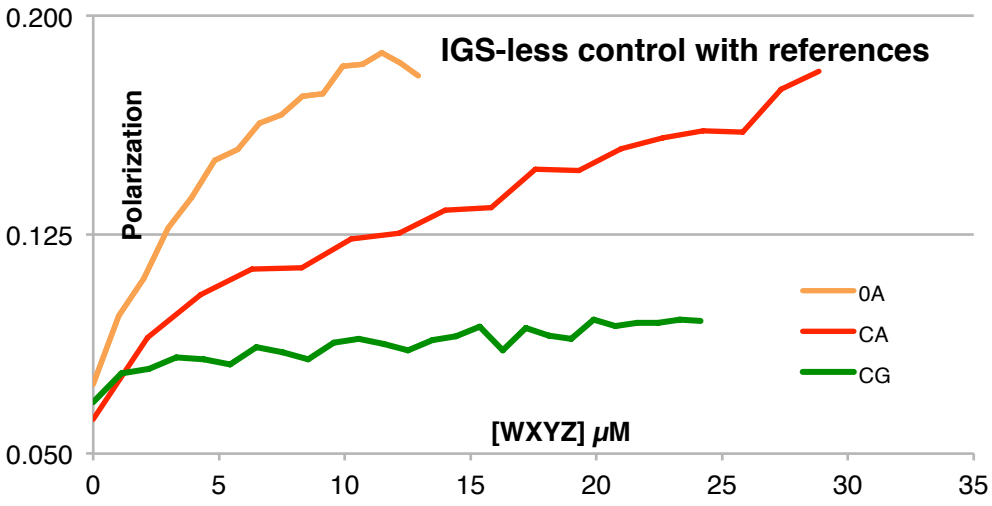
Figure 8a: Native gel GGG no duplex at 5 mM 1x reaction buffer



µM 0.7 2.0 3.3 4.7 6.0 7.3 8.6 10.0 11.3 12.5 14.1 15.3

To provide a control for all 16 genotype combination experiments, a ribozyme lacking an IGS was prepared to observe in the presence of a duplex. Since the binding

Figure 9: IGS-less ribozyme/duplex pairing



site has effectively been removed from the ribozyme, no specific IGS/tag binding interactions can occur. The result of an IGS-less experiment is shown in Figure 9, with two other previously shown experiments added for scaling. The distinct leveling-off seen in Figures 9 and 13e provide evidence that there are indeed non-specific, tertiary binding interactions occurring with the ribozyme/duplex complex. As a control reaction, this experiment provides a standardized model of the amount of non-specific and/or tertiary binding interactions that are presumably occurring with all other genotype combinations. Thus it is hypothesized that any genotype combination that has a lower K_d than the K_d of the IGS-less reaction has additional favorable binding interactions that further lower the K_d value. On the other hand, any combination that has a higher K_d than the IGS-less reaction must have a higher K_d due to some sort interference of the observed non-specific/tertiary binding interactions.

While Figure 6a-d offers useful preliminary information about the behavior of various binding interactions, the data must be fitted to a binding plot curve in order to determine the rate constant K_a. ($1/ K_a$) gives K_d, the dissociation constant in units of μM, which offers a tangible representation of how good or poor of a binding pair a particular genotype combination is. K_d indicates the required concentration of ribozyme present for half of the duplex present to be bound. A better combination such as a Watson-Crick pairing should thus have a lower K_d value than a poor combination such as AA, because a more stable canonical bond should require a lower ribozyme concentration for half of the ribozyme to be bound, since the canonical bonds are more favorably formed than those such as AA.

From the determined average polarization value for each 1-2 μL addition of ribozyme, the anisotropy value can be determined from the formula A=(2P)/(3-P). The anisotropy value is plotted against ribozyme concentration to generate binding plots, using the formula y=m1+(m2-m1)*(m3*M0/(1+m3*M0)); where m1=anisotropy value when the concentration of ribozyme is zero, m2=estimated maximum anisotropy, and m3 is an estimate of what the K_a will be for the particular genotype combination. M0 is the total ribozyme concentration. Figures 10a-13e show binding plot curves for all 16 genotype combinations, plus a GC experiment with reduced 5mM Mg^{2+} reaction buffer (12e), and an IGS-less ribozyme with a duplex (13e). The K_a and K_d values obtained

from each plot have been tabulated in Table 14. In general, the Watson-Crick combinations at the top of Table 5 should exhibit the most exaggerated leveling-off in Figures 10-13, followed by intermediate combinations in the middle of the table. The poorest combinations at the bottom of Table 5 are expected to have little to no curvature in their respective binding plot.

Figure 10a-d: Binding plots of 'A-' ribozyme. Note that figures 10a-13e are so ordered from left to right, top to bottom.

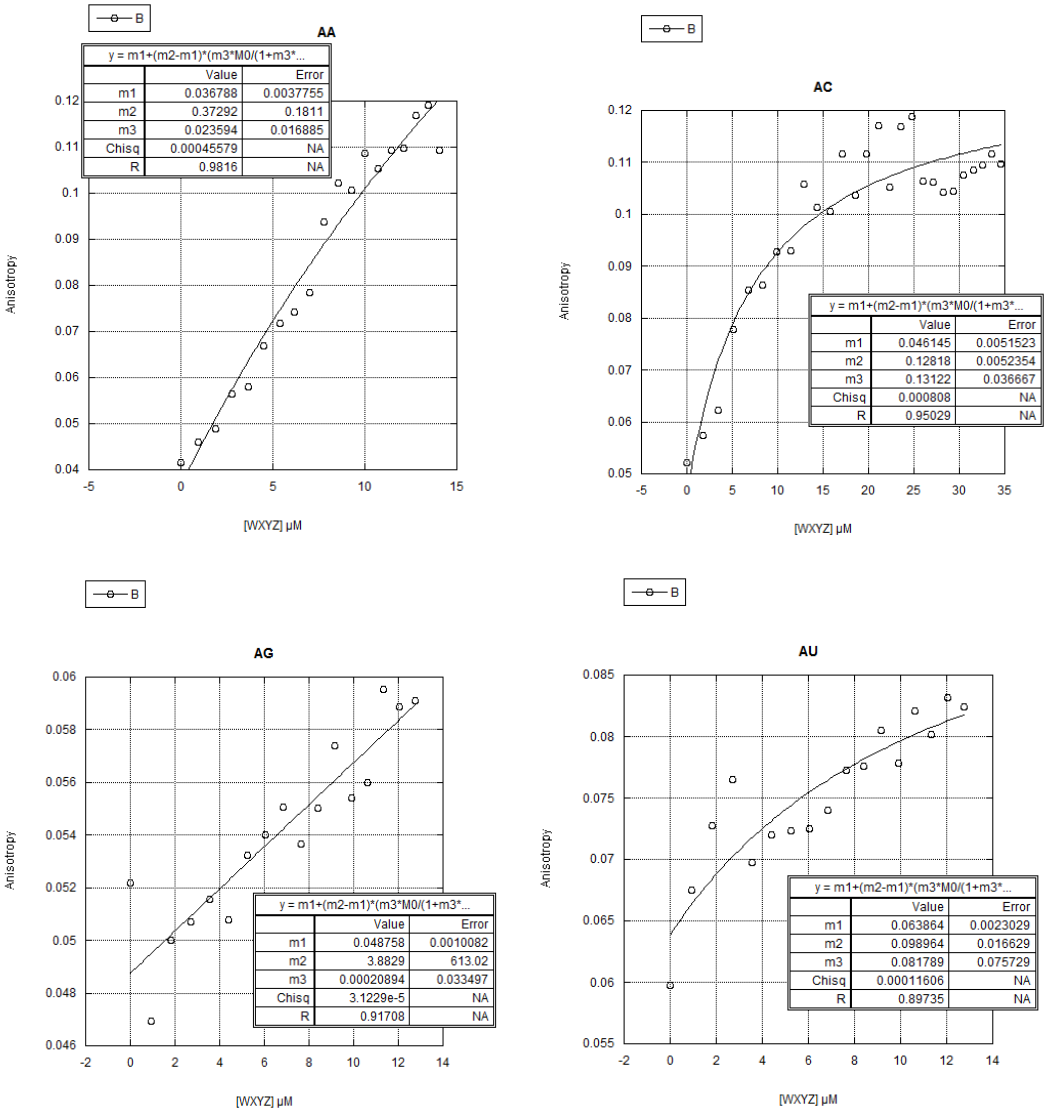


Figure 11a-d: Binding plots of ‘C-‘ ribozyme

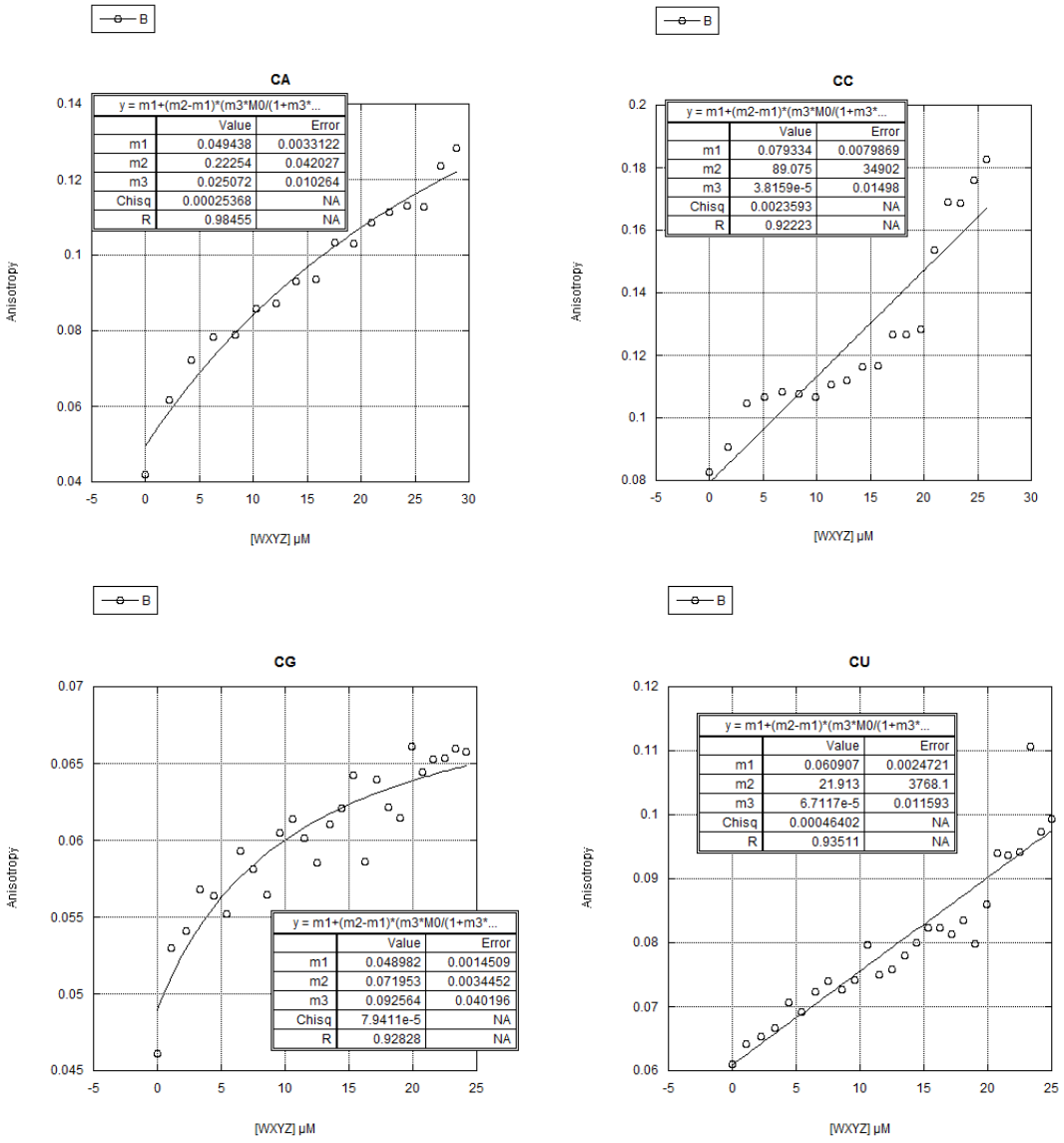
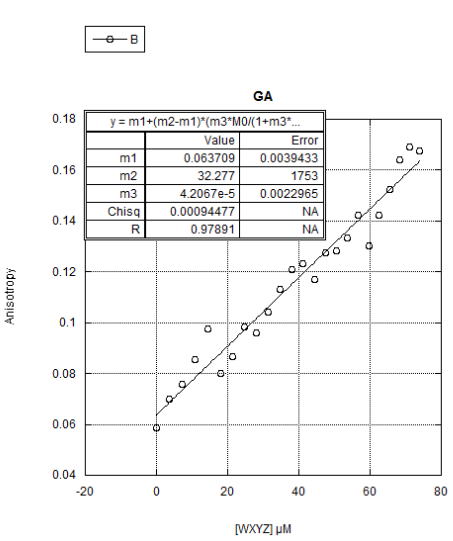
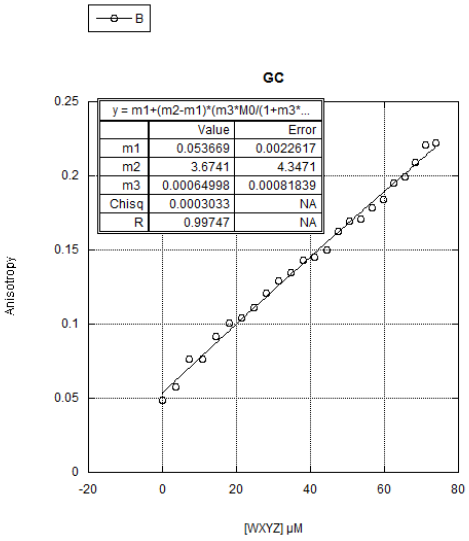
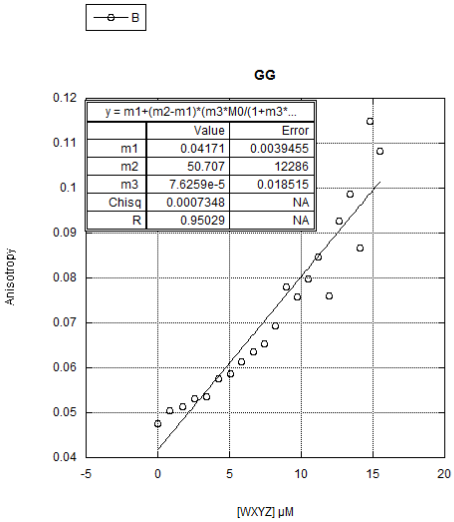
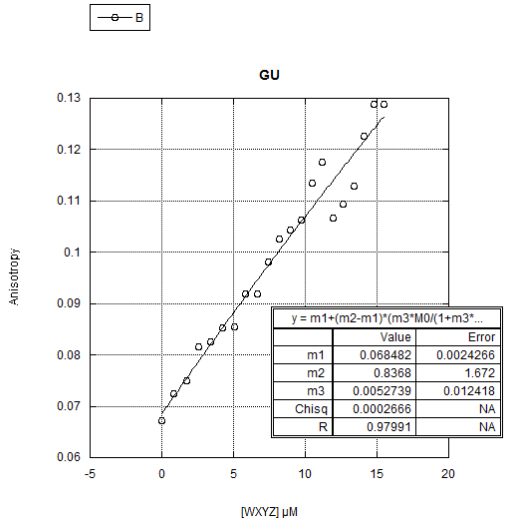


Figure 12a-e: Binding plots of 'G-' ribozyme









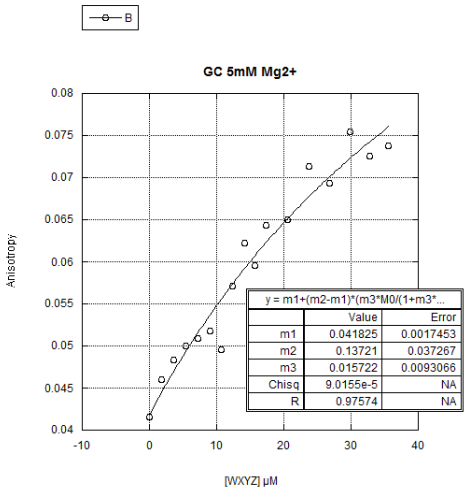


Figure 13a-e: Binding plots of 'U-' ribozyme

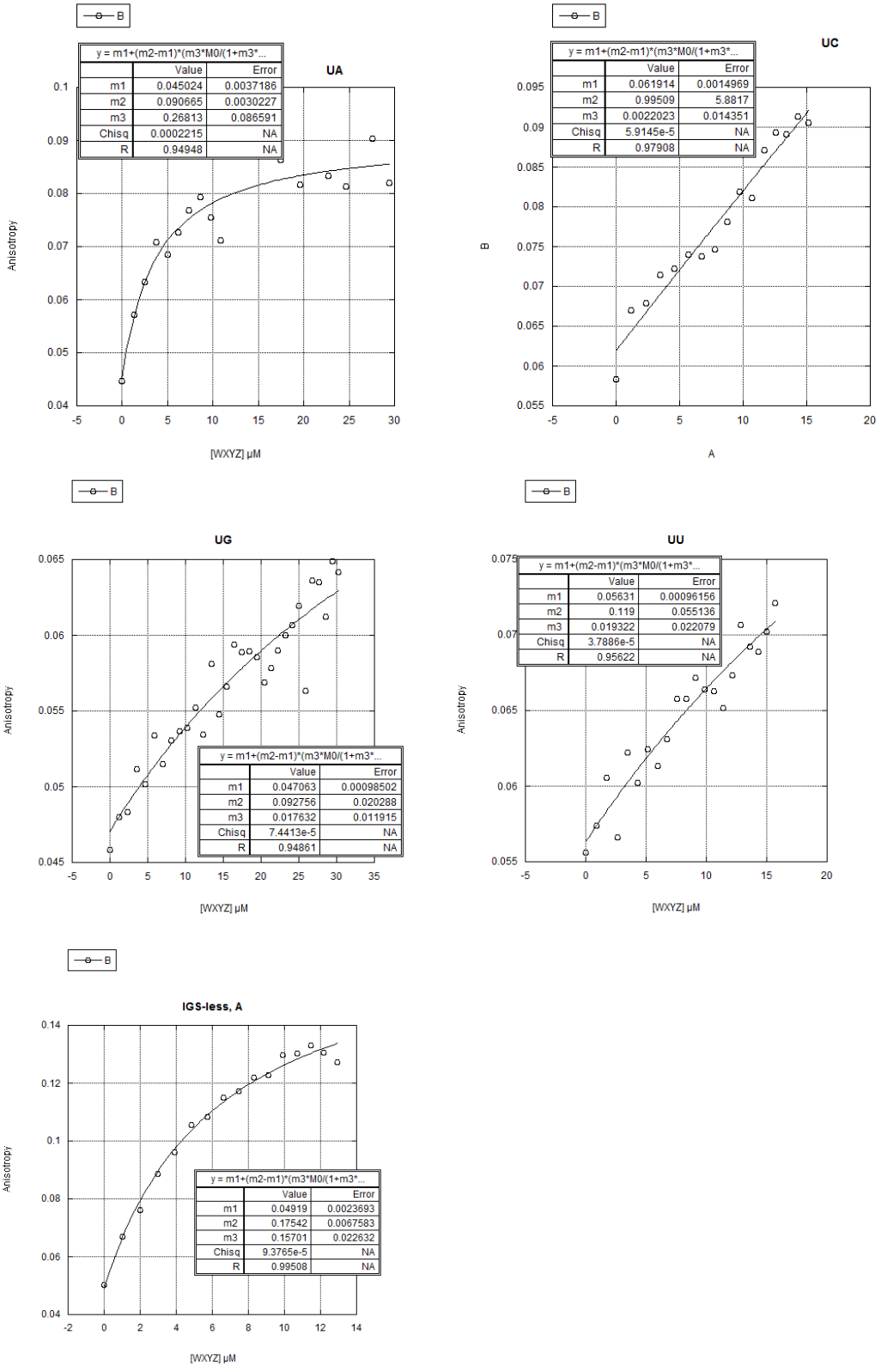


Table 14: Experimentally determined dissociation constants and calculated Gibbs energies of dissociation

genotype combination	experimental k_a (min^{-1})	experimental K_d (μM)	ΔG°48 (kcal/mol)
CG	0.0415	10.8	7.29
AU	0.0319	12.2	7.21
UA	0.0197	3.73	7.97
GC	0.0125	1.54×10^{3}	4.13
GU	0.0091	1.89×10^{2}	5.47
AC	0.0069	7.62	7.52
UG	0.0049	56.7	6.24
UC	0.0038	4.54×10^{2}	4.91
UU	0.0022	51.8	6.29
CA	0.0020	39.9	6.46
CC	0.0016	2.62×10^{4}	2.32
GG	0.0006	1.31×10^{4}	2.76
GA	0.0005	2.38×10^{4}	2.39
AA	0.0004	42.4	6.42
CU	0.0004	2.16×10^{3}	3.92
AG	0.0001	4.79×10^{3}	3.41
GC (5 mM Mg^{2+})	N/A	63.6	6.16
⃠A	N/A	6.37	7.63

Table 14 has the 16 genotype combinations listed in the same order as in Table 5 for comparative purposes. Recall that it is not expected that the predicted rate constants (k_a) will be the same as the experimental association constant (K_a) as these are not the same constant. Table 14 shows experimentally determined K_d values in the third column; where K_a can be easily inferred from finding $1/K_d$. Thus if there is a strong correlation between 'MN' combination and thermodynamic strength of that bond as is hypothesized, then K_d should behave inversely as k_a, meaning as k_a decreases K_d should increase. Indeed there does seem to be some sort of loose correlation in increasing value of K_d as the binding combination becomes poorer, however there are several significant deviations from this correlation. Also, since the Gibbs free energy values in Table 5 are the energy of hybridization (association) of the 3-nucleotide active site while the Gibbs free energy

in Table 14 is of dissociation, we also expect to see an inverse correlation between these values.

It has already been shown that combinations with a GGG ribozyme (GG, GC, GA, GU) will have ribozyme aggregation at high concentrations using the standard 100mM Mg^{2+} reaction buffer. This aggregation behavior accounts for the experimental K_a for GC being significantly lesser than the other Watson Crick combinations (CG, UA, AU). Note that the 5mM Mg^{2+} GC experiment in Table 13 yielded a K_a value quite similar to CG, UA, and AU. Aside from combinations that include the GGG ribozyme, there are other instances of a break in the decreasing trend, such as AC, CA and AA. The Gibbs free energy of dissociation for each combination is in the far right column, calculated from the combination's respective K_d value. The energy values show that dissociation among better binding combinations is less favorable than dissociation among poor binding combinations.

As a testament to the significance of non-specific tertiary binding interactions in the *Azoarcus* group I intron, note that the K_d value for the '0A' (IGS-less ribozyme with 'CAU' duplex) is 6.37 μM. There is only one genotype combination in Table 14 that has an experimental K_d that is less than this; UA at 3.73 μM. Even accounting for error in the experimentally determined K_d values, this is significant because it indicates that almost all genotype combinations are experiencing some degree of binding interference because of the genotype combination. In other words, the concentration at which half of all duplex molecules are bound to ribozyme is being increased possibly because of tertiary interactions becoming hindered due to the specific genotype combination being an unfavorable binding event. In this study, only the primary interaction between the two nucleotides that comprise the genotype combination is truly studied, while there is strong evidence for non-specific and tertiary molecular interactions occurring as well.

In order to quantitatively determine whether there is a statistically significant correlation between the experimental rate constants and equilibrium constants of the genotype combinations, a Spearman's rank order correlation test was performed. All four genotype combinations that included the GGG ribozyme were removed from consideration, as it has been shown that these molecules behave in an abnormal manner, thus skewing their experimental K_d values. A rank-order comparison of the remaining 12

combinations yielded a correlation coefficient of -0.743, which is considered to be statistically significant.[12] This indicates that while there is a distinguishable correlation between these values, the correlation itself is weak. This leads to the conclusion that the 'MN' combination is not the sole determinant of the thermodynamic strength of the bond, as the correlation coefficient would be stronger and thus closer to -1 if it were. A Spearman's test was also performed between the predicted Gibbs energy values in Table 5 and the determined Gibbs values in Table 14, with the same 12 combinations. This test yielded a correlation coefficient of -0.458, indicating the two sets are not significantly associated. This reinforces the knowledge that the predicted values in Table 5 do not accurately represent the entirety of the ribozyme as the values only reflect the energy associated with hybridization between the IGS and tag.

Conclusion

In conclusion, all possible genotype combinations between a group I intron with the IGS 'GMG' and a duplex with the tag 'CNU' were observed using fluorescence anisotropy to determine the equilibrium constants of association (K_a) and dissociation (K_d), and Gibbs free energy of dissociation. This modified reaction was designed to mimic the recombination reactions of the *Azoarcus* group I intron, where the intron molecule can assemble itself from four fragments using recombination reactions in an autocatalytic fashion such that an assembled ribozyme can catalyze more recombination reactions. The experimentally determined equilibrium constants were found to be weakly correlated to the rate constants for each respective genotype combination by Spearman's rank order correlation test.

This research has provided evidence for the hypothesis that there is a correlation between genotype combination and catalytic activity with the thermodynamic strength of the bond. It was shown that in general, canonical combinations exhibit a lower K_d than intermediate combinations, which exhibit a lower K_d than poorer genotype combinations. It is concluded that while there is a significant correlation between the predicted rate constants of the recombination reaction and the dissociation constants of the modified reaction that several discrepancies in the correlation allude to higher-order molecular interactions occurring. The analysis of a ribozyme without an IGS yielding a low K_d value confirms that tertiary interactions play a significant role in the ultimate K_d and Gibbs free energy of each genotype combination. In future research, these possibilities of other molecular interactions will be addressed.

References

1. Hayden, Eric J., Lehman, Niles (2006). "Self-Assembly of a Group I Intron from Inactive Oligonucleotide Fragments." *Chemistry & Biology* **8:** 909-18.

2. "Alternative Definitions of Life: Perspective Matters." *NASA's Alternative Definition of Life,* Simon Fraser University, n.d. Web. 20 May 2016.

3. Joyce, Gerald F. (2002). "The Antiquity of RNA-based Evolution." Figure 1: "Timeline of events pertaining to the early history of life on Earth, with approximate dates in billions of years before the present." *Nature* **418:** 214-21.

4. "Gene Expression." Figure: "Structural differences between deoxyribose and ribose." *BIO 202 Genetics: From Mendel to the Human Genome Project.* University of New Mexico, 2013. Web. 20 May 2016.

5. Vaidya, Nilesh *et al.* (2012). "Spontaneous Network Formation Among Cooperative RNA Replicators." *Nature* **491:** 72-77.

6. Myerson, Roger B. (1991). *Game Theory: Analysis of Conflict,*Harvard University Press, p. 1. Chapter-preview links, pp. vii–xi.

7. Hayden *et al.* (2008). "Systems Chemistry on Ribozyme Self- Construction: Evidence for Anabolic Autocatalysis in a Recombination Network." *Angewandte Chemie International Edition* **47:** 8424–28

8. Yeates *et al.* (2016). "Dynamics of prebiotic RNA reproduction illuminated by chemical game theory." *Proceedings of the National Academy of Sciences* **18:** 5030-35.

9. von Kiedrowski, G (1986). "A Self-Replicating Hexadeoxynucleotide." *Angewandte Chemie International Edition* **25**: 932-35.

10. Yeates, J.M. (2015). *The Foundations of Network Dynamics in an RNA Recombinase System* (Doctoral dissertation). Portland State University Department of Chemistry.

11. Gleitsman, K.R., Herschlag, D.H. (2014). "A kinetic and thermodynamic framework for the *Azoarcus* group I ribozyme reaction." *RNA* **20**: 1732–46.

12. "Spearman's Rho Calculator." *Spearman's Rho Calculator (Correlation Coefficient).* Jeremy Stangroom, 2016. Web. 28 May 2016.